Ensemble of SLR systems for forensic evidence

Federico Veneri, MS ; Danica Ommen, Phd.

Iowa State University – Department of Statistics

Iowa State University – Center for Statistics and Application in Forensic Evidence

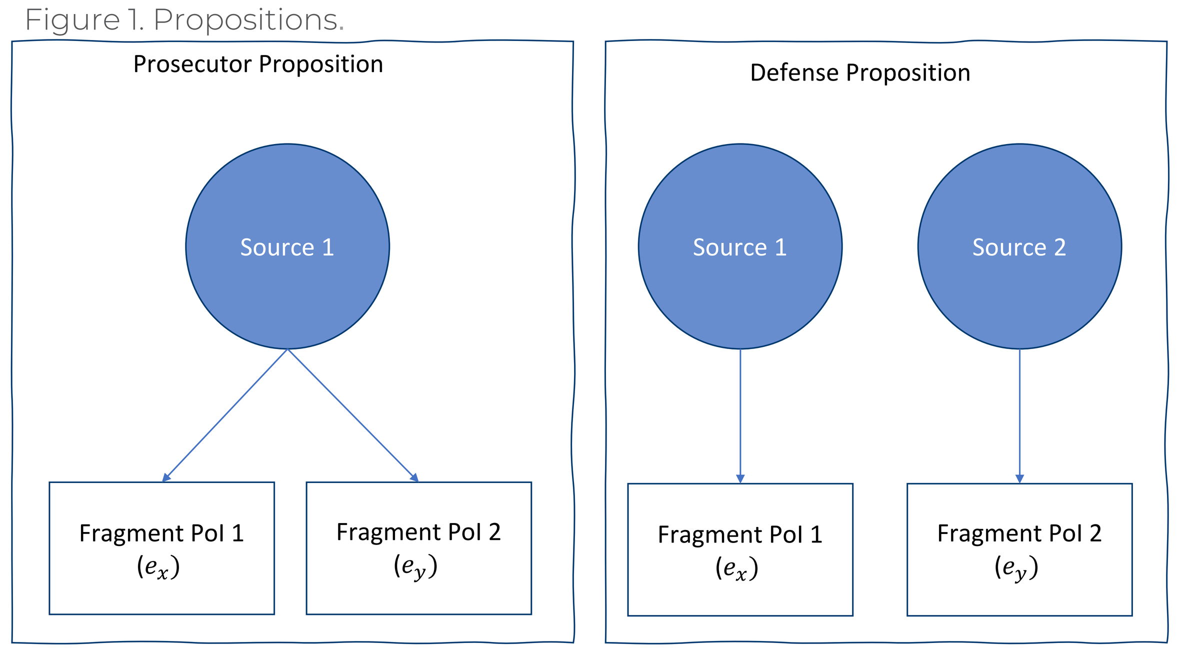
1. Objectives

- Score likelihood ratios (SLR) are an alternative way to provide a numerical assessment of evidential strength when contrasting two propositions.
- The SLR approach focuses on a lower-dimensional (dis)similarity metric and avoids distributional assumptions regarding the features [1].
- SLR can be developed in two steps:
  1. Constructing comparison metric: $C(.)$
  2. Estimating the distribution of the metric under both propositions: $g(C(.)|H_p)$ and $g(C(.)|H_D)$ or directly estimating the ratio $g(C(.)|H_p)\,/g(C(.)|H_D)$
- Yet, the independence assumption needed is not met and is often overlooked.
- We introduce an ensemble approach to remedy this lack of independence and improve SLR performance.
- To illustrate our approach, we use real forensic glass data [2] under the common source problem [3].
- We use a Random Forest (RF) based score as a comparison metric, and for estimating the density ratio, we used a logistic classifier.

2. The Independence assumption in SLR.

- Consider a hypothetical common source problem with two individuals (PoI), and one glass fragment recovered from each.
- The forensic question: do these fragments share an unknown common source?

Figure 1. Propositions.



Available Data:

$$E = \{u_x, u_y\}, I = \{A\}$$

- u_x and u_y measurements from evidence recovered.
- A background measurements.

$$A = \{A_{ij}: i^{th}\ source, j^{th}\ fragments\}$$

- From A pairwise comparisons are constructed and classified as known match (KM) or known non match (KNM).

- The forensic proposition can be translated into sampling models that generated the data M_p and M_d respectively, and training/estimation set defined [3,4].

Under M_p, KM are used:

$$C_{CS_P} = \{C(A_{ij}, A_{kl}):\ i = k\}$$

The set of pairwise comparisons of fragments with the same source.

Under M_d, KNM are used:

$$C_{CS_D} = \{C(A_{ij}, A_{kl}):\ i \neq k\}$$

The set of pairwise comparisons between two different sources.

- Machine learning-based comparison metrics and density estimation procedures rely on the independence assumption, but this assumption is not met:
- At a source level: Sources enter comparisons multiple times.
  - In C_{cs_P} : multiple within comparisons for the same source.
  - In C_{cs_D} : multiple between comparison use same sources.
  - Same sources appear in both C_{cs_P} and C_{cs_D}
- At an item level: glass fragments enter comparison multiple times.
  - e.g., $C(A_{11}, A_{21}), C(A_{11}, A_{31})$

3. Sampling and ensembling.

- We propose a sampling approach to construct independent datasets.

Strong Source Sampling Algorithm (SSSA)

1. Construct all pairwise comparisons available.
2. For KM pairs:
   Sample randomly one pair to be used in the final database.
   Remove all pairs in the dataset involving sources selected in the previous step.
3. For KNM
   Sample randomly one pair to be used in the final database.
   Remove all pairs in the dataset involving sources selected in the previous step.
4. Repeat 2 and 3 until data is exhausted.

Result: A pairwise database where sources are not compared multiple times.

- The sampling approach reduces the sample size drastically but allows for ensembling to improve the SLRs

Ensembling Score likelihood System (ESLR)

1. For 1:M
2. Use SSSA to generate a pseudo training set.
3. Train a machine learning comparison metric.
4. Use SSSA to generate a pseudo estimation set.
5. Predict a comparison score for all cases on the estimation set.
6. Estimate the distribution of scores under both propositions (or ratio estimator).
7. Store the comparison metric and estimated distribution

Result: M-SLR system (M dyads of RF - ratio estimator) that can be combined into a final ensembled SLR.

4. Application.

- Using the validation set from [2] and SSSA, we created a balanced hold out a set of 106 KM and 106 KNM independent pairs.
- Using the primary dataset from [2], we ran M=50 iteration of the ESLR algorithm and aggregated the final scores using the median.
- Figure 2 shows by SLR system (x-axis) the evidential value in a log 10 scale (y-axis) for each hold-out pair represented by connected points. Figure 3 present the partial ensemble, meaning how the SLR evolves as an additional SLR system is incorporated.

Figure 2. Test set pairs under individual SLRs

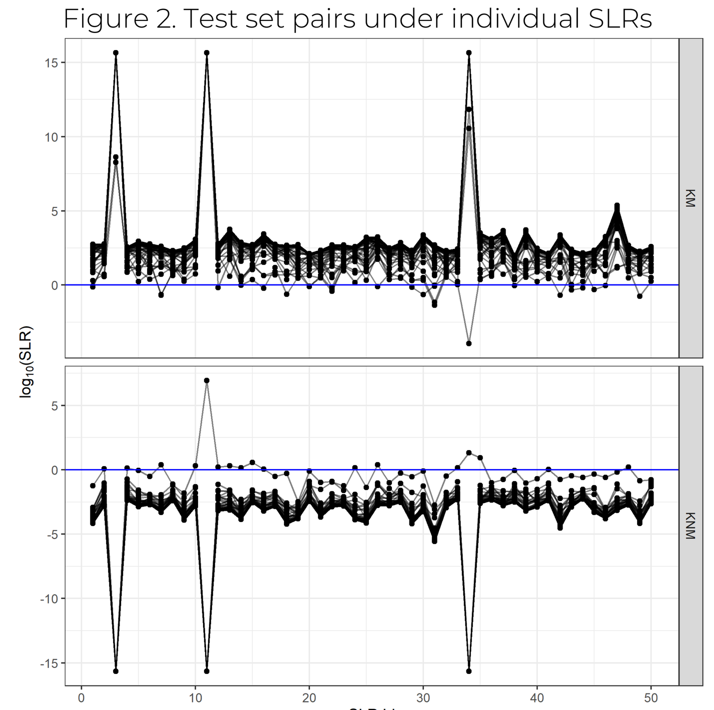
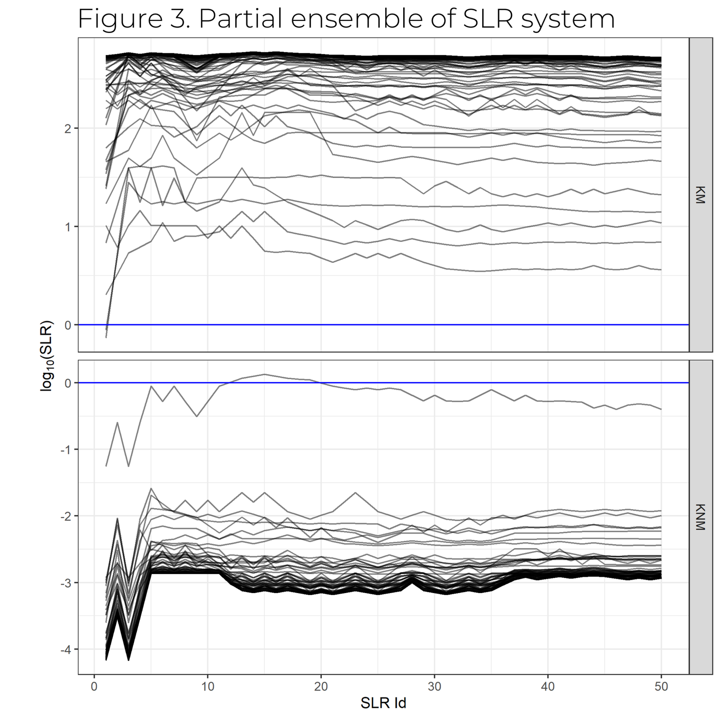


Figure 3. Partial ensemble of SLR system



- Individual SLR may provide evidence in the wrong direction. However, by combining different SLRs we can provide more decisive conclusions in the correct direction, improving the performance of the final SLR scores.

5. Conclusions and future work.

- Independence is a crucial assumption for machine learning and density estimation, cornerstones in machine learning-based score likelihood ratios.
- We introduce a sampling algorithm that remedies the dependence structure in forensic comparisons.
- Over these new data sets, standard estimation techniques that require independence can be applied.
- As in ensemble learning, multiple SLR systems can be developed, allowing them to learn from a partial view of the data and aggregate their conclusion into a final SLR score.
- We are conducting further studies to compare this approach to traditional SLR and LR using data from different forensic domains.
- Currently, aggregation is done naively, but more sophisticated procedures are being tested, like weighting the SLRs system according to their performance.

6. References.

1. Hal S Stern, "Statistical Issues in Forensic Science," Annual Review of Statistics and Its Application 4 (2017): 225–244.
2. van Es, A., Wiarda, W., Hordijk, M., Alberink, I., & Vergeer, P. (2017). Implementation and assessment of a likelihood ratio approach for the evaluation of LA-ICP-MS evidence in forensic glass analysis. Science & Justice, 57(3), 181-192.4.
3. Ommen, D. M., & Saunders, C. P. (2018). Building a unified statistical framework for the forensic identification of source problems. Law, Probability and Risk, 17(2), 179-197.
4. Johnson, M. Q., & Ommen, D. M. (2021). Handwriting identification using random forests and score-based likelihood ratios. Statistical Analysis and Data Mining: The ASA Data Science Journal.

7. Acknowledgments

The data (van Es et al. 2017) was kindly supplied by the Netherlands Forensic Institute (NFI), but the NFI was not further involved in this research. The data are available upon request by emailing p.zoon@nfi.nl.

This work was funded (or partially funded) by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreements 70NANB15H176 and 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, Duke University, University of California Irvine, University of Virginia, West Virginia University, University of Pennsylvania, Swarthmore College and University of Nebraska, Lincoln.